

# Pole Frequency Estimation of Speech by a Restricted Linear Prediction Method

メタデータ	言語: jpn 出版者: 公開日: 2017-10-02 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	<a href="http://hdl.handle.net/2297/3351">http://hdl.handle.net/2297/3351</a>

## 拘束線形予測分析による音声の極周波数推定

正員 三好 義昭<sup>†</sup> 正員 角所 取<sup>††</sup>

Pole Frequency Estimation of Speech by a Restricted Linear Prediction Method

Yoshiaki MIYOSHI<sup>†</sup> and Osamu KAKUSHO<sup>††</sup>, Members<sup>†</sup> 金沢大学教育学部, 金沢市

Faculty of Education, Kanazawa University, Kanazawa-shi 920 Japan

<sup>††</sup> 龍谷大学理工学部, 大津市

Faculty of Science and Technology, Ryukoku University, Otsu-shi, 520-21 Japan

あらまし 極の位置を  $z$  平面の単位円上に拘束することにより, 極の周波数のみを推定する拘束線形予測法の定式化を行い, 本方法の正規方程式の次元ならびに高次方程式の次数が通常の線形予測法の 1/2 に半減することを示す。そして, 合成音の極周波数推定に適用し, 本方法の特徴ならびに有効性を実験的に示す。

## 1. まえがき

今日, 音声の重要な特徴パラメータであるホルマントの推定手法として, 線形予測分析が広く活用されている。しかしながら, 通常の線形予測分析では, ホルマント帯域幅の正確な推定が一般に困難であり, 音声認識におけるホルマント帯域幅の有用性はまだ検討段階でもあることから, 線形予測分析により得られる帯域幅の情報はホルマントに対応する極の選定程度にしか使用されていないのが現状である。

本論文では, 極の周波数のみを推定する一手段として, 極の位置を  $z$  平面の単位円上に拘束した拘束線形予測法の定式化を行い, 本方法を合成音の極周波数推定に適用し, そのホルマント周波数推定精度の検討ならびに線スペクトル対 (LSP)<sup>(1)</sup> との比較を行った結果について述べる。

## 2. 極の位置を拘束した線形予測分析

通常の線形予測法では, 周知のように, 声道伝達特性  $H(z)$  を,

$$H(z) = \frac{1}{1 + \sum_{k=1}^p a_k z^{-k}} \quad (1)$$

但し,  $a_k$ : 第  $k$  予測係数  
なる全極型で記述できるものと仮定している。ここで, 次数  $p$  が偶数の場合, 式(1)は,

$$H(z) = \frac{1}{\prod_{i=1}^{p/2} (1 + a_i z^{-1} + b_i z^{-2})} \quad (2)$$

と共役複素極の積の形で記述できる。

今,  $1 + a_i z^{-1} + b_i z^{-2} = 0$  の根を  $z_i$  とすると  $|z_i| = \sqrt{b_i}$  より,

$$b_i = 1 \quad (3)$$

但し,  $i=1, 2, 3, \dots, p/2$

のとき,  $H(z)$  の極はすべて  $z$  平面の単位円上に存在する。このとき, 式(2)の分母は,

$$\prod_{i=1}^{p/2} (1 + a_i z^{-1} + z^{-2}) = 1 + c_1 z^{-1} + c_2 z^{-2} + \dots + c_2 z^{-p+2} + c_1 z^{-p+1} + z^{-p} \quad (4)$$

但し,  $c_i = f(a_j)$ ,  $i=1, 2, \dots, p/2$

と,  $z^{-i}$  の係数  $c_i$  と  $z^{-p+i}$  の係数  $c_{p-i}$  が等しくなることより, 式(1)において,

$$a_i = a_{p-i} \quad (5)$$

但し,  $a_0 = 1$ ,  $i=0, 1, 2, \dots, p/2$

となれば, すなわち,  $\{a_i\}$ ,  $i=1, 2, 3, \dots, p$  が  $p/2$  番目の予測係数  $a_{p/2}$  を中心とした対称形となれば, 声道伝達特性  $H(z)$  の極はすべて  $z$  平面の単位円上に存在することになる。このとき,

$$1 + \sum_{k=1}^p a_k z^{-k} = 0 \quad (6)$$

は式(5)より相反方程式となり, 式(6)を解くことは  $x = z + z^{-1}$  に関する  $p/2$  方程式を解くことと,  $z$  に関する 2 次方程式  $z^2 - xz + 1 = 0$  を解くことに帰着する<sup>(2)</sup>。そして, 予測係数  $\{a_i\}$  も式(5)のもとで, 分析フレーム内の  $N-p$  個の音声標本値  $\{s_n\}$ ,  $n=p+1, p+2, \dots, N$  に対する予測誤差の 2 乗平均最小の条件より,

$$\sum_{k=1}^{p/2} \beta_k \sum_{n=p+1}^N (s_{n-k} + s_{n+k-p})(s_{n-i} + s_{n+i-p}) = - \sum_{n=p+1}^N (s_n + s_{n-p})(s_{n-i} + s_{n+i-p}) \quad (7)$$

$$\alpha_j = \beta_j, \quad \alpha_{p/2} = 2\beta_{p/2}$$

但し,  $i=1, 2, 3, \dots, p/2$

$$j=1, 2, 3, \dots, p/2-1$$

なる連立 1 次方程式の解として,  $p/2$  個の予測係数  $\{a_1, a_2, \dots, a_{p/2}\}$  のみを求めればよいことになる。

以上のように, 極の位置を  $z$  平面の単位円上に拘束することにより, 極の帯域幅の情報は得られないが, 正規方程式の次元ならびに高次方程式の次数をそれぞれ半減することができる。この拘束がホルマント周波数推定精度にどの様に影響するかを解析的に明らかにするのは困難であるため, 以下, 合成音によるシミュレーションにより, その検証を行う。

## 3. 合成音によるシミュレーション結果

## 3.1 ホルマント周波数推定精度

標準化周波数 10 kHz, 励振源: ピッチ周期 8 ms の

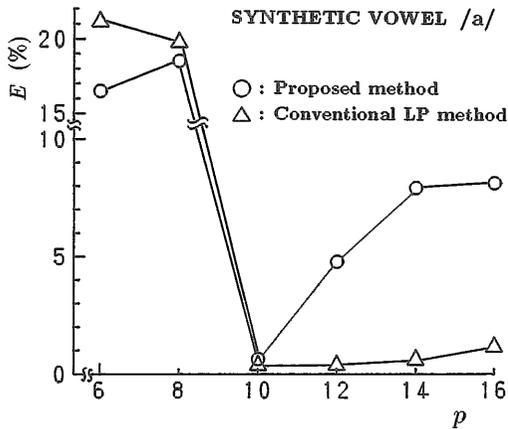


図1 ホルマント周波数推定誤差の分析次数  $p$  依存性  
Fig.1 The  $p$ -dependency of the formant estimation error  $E$ .

Rosenberg 波<sup>(3)</sup>, ホルマント周波数:  $F_1=812.5$  Hz,  $F_2=1312.5$  Hz,  $F_3=2687.5$  Hz,  $F_4=3437.5$  Hz,  $F_5=4437.5$  Hz, 放射特性: 6 dB/oct として作成した合成母音/a/における第1～第3ホルマント周波数推定誤差  $E$  の分析次数  $p$  依存性を図1に示す。但し, 前処理として1階差分後, 分析窓長  $T_a=25.6$  ms, フレームシフト間隔 0.2 ms で1周期に渡って分析した計40フレームの平均値である。そして, ○印: 本方法, △印: 通常の線形予測法 (共分散法) の結果である。

今の場合, 上記5個のホルマントを用いて声道特性を設定しているので, 声道伝達特性の次数  $p_0=10$  となる。従って, 図1に示されているように, 分析次数  $p < 10 (=p_0)$  では, ホルマント周波数推定誤差は両方法とも大きくなるが,  $p=10 (=p_0)$  の場合, 通常の線形予測分析による誤差は0.4%であるのに対して, 本方法による誤差は0.7%と若干劣るが, 本方法でも通常の線形予測分析とほぼ同等の精度でホルマント周波数が精度良く推定可能であると言える。しかしながら, 通常の線形予測分析は  $p \geq 14$  においてホルマント周波数推定誤差が若干増大してはいるが,  $p \geq p_0$  であれば, ホルマント周波数推定誤差に及ぼす分析次数依存性はほとんどないのに対して, 本方法は  $p > 10$  において, ホルマント周波数推定誤差が分析次数に依存して増大し, 通常の線形予測分析とは大きく異なった特性となる。すなわち,  $p > p_0$  の場合には, ホルマントに対応しない極, いわゆる擬似ホルマントが生じるが, 通常の線形予測分析では擬似ホルマントの帯域幅が一般に大きいため, ホルマントに対応する極への影響が小さく,

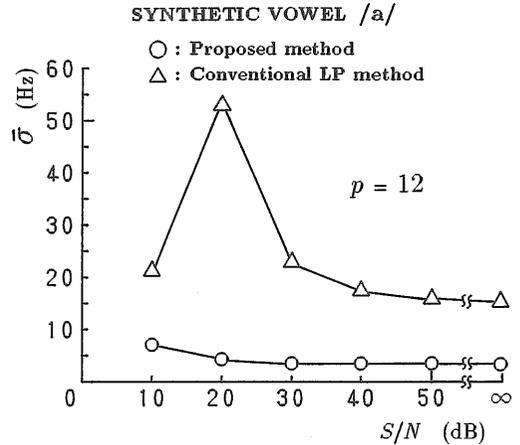


図2 極周波数推定値の平均標準偏差のSN比依存性  
Fig.2 The  $S/N$ -dependency of the average standard deviation of estimated pole frequencies.

分析次数を  $p > p_0$  としても, ホルマント周波数推定精度が急激に悪くなることはないのに対し, 本方法ではすべての極の位置が  $z$  平面の単位円上に拘束されているため, 擬似ホルマントの存在が大きく影響し,  $p > p_0$  となるとホルマント周波数推定精度が急激に悪くなると言える。

### 3.2 極周波数推定値の頑健性

通常の線形予測分析において分析次数  $p$  を  $p > p_0$  とした場合, 母音定常部においても分析位置により, 特に擬似ホルマントの位置が大きく変動するため, 得られる極のすべてを音声の特徴パラメータとして利用することができず, ホルマントに対応する極の選定が必要となる<sup>(4)</sup>。一方, 本方法は極の位置が  $z$  平面の単位円上に拘束されていることより, 定性的には推定極の位置の変動が予測誤差に大きく反映するため, 予測誤差の2乗平均最小の条件より, 母音定常部のようなホルマント周波数がほぼ一定とみなせる音声区間においては,  $p > p_0$  の場合でもホルマントに対応しない極を含めて極周波数推定値が分析位置等の影響で大きく変動することはないと考えられる。

前述の合成母音/a/において,  $p$  次分析により得られる  $p/2$  個の極周波数各々の標準偏差を平均した式(8)で定義する極周波数推定値の平均標準偏差  $\bar{\sigma}$  のSN比依存性を図2に示す。但し, 分析次数  $p=12$  とし, その他の分析条件ならびに図中の記号等は図1と同じである。なお, 極の対応関係も含めた変動度合を評価するため, 式(8)の  $F_{ij}$  は第  $j$  分析フレームで得られる極周波数を単に周波数の小さい順に番号付けしたも

のである。そして、通常の線形予測分析の  $\sigma$  算出には極周波数推定値が 0 Hz または 5000 Hz となる分析フレーム ( $S/N=20$  dB では 7 フレーム,  $S/N=30, 40, 50$  dB および  $\infty$  ではそれぞれ各 1 フレーム) は除外したので、例えば、 $S/N=20$  dB での式(8)の分析フレーム数は  $M=33$  と他の SN 比より若干少なくなっているが、本方法では、そのような分析フレームは生じなかったため、各 SN 比とも  $M=40$  である。

$$\sigma = \frac{2}{p} \sqrt{\frac{1}{M} \sum_{j=1}^M (F_{ij} - \bar{F}_i)^2} \quad (8)$$

$$\text{但し, } \bar{F}_i = \frac{1}{M} \sum_{j=1}^M F_{ij}$$

$F_{ij}$  : 第  $j$  分析フレームの第  $i$  極周波数

$M$  : 分析フレーム数

$p$  : 分析次数

図 2 より、通常の線形予測分析により得られる極周波数は  $S/N=\infty$  の場合でも分析位置により変動し、平均標準偏差  $\sigma=15.3$  Hz となるのに対し、本方法では 3.4 Hz と極周波数推定値の分析位置依存性が大幅に改善することがわかる。そして、通常の線形予測分析の極周波数推定値の変動は SN 比が減少すると共に増大し、 $S/N=20$  dB において  $\sigma=52.9$  Hz と特に大きくなるが、これは分析位置により擬似ホルマントが、今の場合、第 3～第 4 ホルマント間に生じたり、第 4～第 5 ホルマント間に生じたりするため、いわゆる極周波数推定値の順序関係が乱れているためである。そして、 $S/N=10$  dB ではこのような順序関係の乱れはなく  $\sigma=21.3$  Hz と、 $S/N=30$  dB の 23.0 Hz よりも若干ながらむしろ改善する結果となっている。一方、本方法の極周波数推定値の変動は  $S/N \leq 20$  dB において若干増大しているが、いずれの  $S/N$  においても通常の線形予測分析よりも大幅に改善すると共に、 $30 \text{ dB} \leq S/N \leq 50$  dB では  $\sigma=3.5$  Hz と、 $S/N=\infty$  の 3.4 Hz とほぼ同じ小さな値となっていることより、本方法では順序関係を保持した安定な極周波数が推定できていると言える。このことは、本方法により得られる極のすべてを音声の特徴パラメータとして利用できることを意味し、本方法は分析次数を少し多めに設定した場合でもホルマントの真偽判定が不要な分析法と言える。

### 3.3 スペクトル対 (LSP) との比較

本方法は音声の線スペクトル表現<sup>(1),(5),(6)</sup>の一方式と言える。以下、線スペクトル化方式の代表例と言える周知の線スペクトル対 (LSP) との比較を行う。

LSP は、周知のようにいったん通常の PARCOR 分

析を行い  $p$  個の PARCOR 係数  $\{k_i\}$ ,  $i=1, 2, \dots, p$  を求め、 $k_{p+1}=\pm 1$  とした 2 種類の  $p/2$  次の高次方程式の解より、計  $p$  本の線スペクトルを得ている<sup>(1)</sup>。それに対して、本方法は式(5)の拘束の基に、式(7)に示す  $p/2$  次の正規方程式の解として、 $p/2$  個の予測係数のみを求め、 $p/2$  次の高次方程式の解として、 $p/2$  本の線スペクトルを得ている。

すなわち、LSP は通常の線形予測分析(自己相関法)を行った後に線スペクトル化の操作を行っているのに対し、本方法は線形予測分析(共分散法)をする段階で線スペクトル化を行っているため、2乗誤差最小基準の線スペクトルが得られ、かつ高次方程式の次数のみならず正規方程式の次元も  $p/2$  に半減している点が従来の LSP 法とは大きく異なる。この違いによる両者の性能の差を解析的に明らかにするのは困難であるの

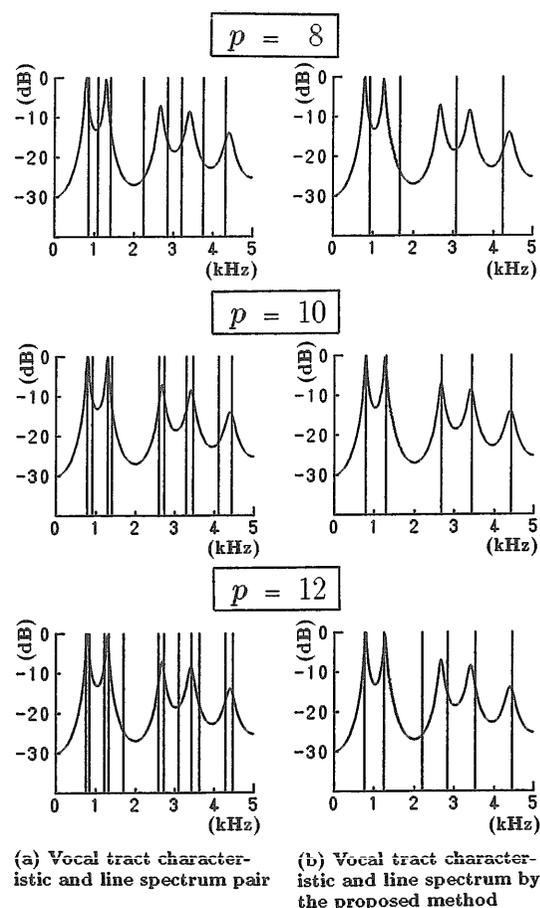


図 3 線スペクトル表現の比較(合成母音/a/)

Fig. 3 A comparison of line spectrum representation for synthetic vowel /a/.

で、以下、合成音の線スペクトル化に適用することにより、主としてホルマント周波数との対応関係の観点から両方式の違いを実験的に示す。

前述の合成母音/a/に適用して得られる線スペクトルの比較を図3に示す。但し、前処理として1階差分後、分析窓長  $T_a=25.6$  ms とし、(a)にLSP法による線スペクトル、(b)に本方法による線スペクトルを示し、図3の上段、中段および下段はそれぞれ分析次数を  $p=8, 10$  および  $12$  とした場合の結果であり、図中のスペクトル包絡は本合成音の声道伝達特性である。

今の場合、前述のように本合成音の声道伝達特性の次数は  $p_0=10$  であるので、分析次数  $p=8 (< p_0)$  の場合、図3(a)、(b)上段の結果より、両方法とも得られる線スペクトルとホルマント周波数との間には特に明確な対応関係は見られないと言える。

分析次数  $p=10 (= p_0)$  の場合、図3(a)、(b)中段の結果より、LSPでは周知のように線スペクトルがホルマントを挟む形で互いに接近して、ホルマントの位置を間接的に示しているのに対して、本方法では  $5 (= p/2)$  本の線スペクトルが得られ、各線スペクトルはホルマント周波数にほぼ一致し、ホルマントの位置を直接的に示していると言える。

分析次数  $p=12 (> p_0)$  の場合、図3(a)、(b)下段の結果より、LSPでは  $1.7$  kHz と  $3.1$  kHz 付近にホルマントに対応しない線スペクトルが生じるが、他の線スペクトルはホルマントを挟む形で互いに接近した1対の組を形成しているのに対して、本方法では  $2.2$  kHz 付近のホルマントに対応しない線スペクトルの影響により、特に第3、第4ホルマントに対応すると考えられる線スペクトルがホルマント周波数に対してそれぞれ高域にシフトしたものとなることがわかる。

#### 4. むすび

極の位置を  $z$  平面の単位円上に拘束した拘束線形

予測法の定式化を行い、本方法のホルマント周波数推定精度ならびに線スペクトル対との比較を合成音により検討した。その結果、本方法は正規方程式の次元ならびに高次方程式の次数を通常の線形予測法の  $1/2$  に半減することができ、かつ分析次数を適切に設定する必要があるが、通常の線形予測法とほぼ同等の精度でホルマント周波数推定が可能であること、ならびに線スペクトル対と比較してホルマントとの対応関係がより明確な線スペクトルが得られることが明らかとなった。更に、極の位置の拘束が極周波数推定値の頑健性にも大きく寄与することが実験的に示された。このことは音声のホルマント構造の情報を担った安定した線スペクトルが得られることを意味し、本方法により得られる線スペクトルは音声認識等における有効な特徴パラメータとなり得ると思われるが、この点については現在検討中である。

#### 文 献

- (1) 菅村 昇, 板倉文忠: “線スペクトル対 (LSP) 音声分析合成方式による音声情報圧縮”, 信学論(A), **J64-A**, 8, pp. 599-606 (1981-08).
- (2) 日本数学会編: “岩波数学辞典(第3版)”, p. 705, 岩波書店 (1990).
- (3) Rosenberg A. E.: “Effect of glottal pulse shape on the quality of natural vowels”, J. Acoust. Soc. Am., **49**, pp. 583-590 (1971).
- (4) 粕谷英樹, 和田充弘, 岡田美智男: “線形予測分析法で得られる極周波数からのホルマント周波数選択アルゴリズム”, 信学論(A), **J66-A**, 11, pp. 1144-1145 (1983-11).
- (5) 嵯峨山茂樹, 板倉文忠: “複合正弦波モデルによる音声スペクトルの分析”, 信学論(A), **J64-A**, 2, pp. 105-112 (1981-02).
- (6) 菅村 昇, 板倉文忠: “線形予測係数の線スペクトル表現とその統計的性質”, 信学論(A), **J64-A**, 4, pp. 323-330 (1981-04).

(平成3年6月12日, 7月15日再受付)